MCC Guidelines for Transparent, Reproducible, and Ethical Data and Documentation (TREDD)

March 6, 2020

MILLENNIUM
CHALLENGE CORPORATION
UNITED STATES OF AMERICA

Table of Contents

Abstract

The purpose of these guidelines is to (i) set forth the principles and procedures for implementing transparent, reproducible, and ethical data and documentation (TREDD) for data activities funded by the Millennium Challenge Corporation (MCC), and (ii) facilitate MCC's observance of the general principles of the Federal Policy for the Protection of Human Subjects or "Common Rule".

Purpose and Scope

Purpose

The purpose of these guidelines is to (i) set forth the principles and procedures for implementing transparent, reproducible, and ethical data and documentation (TREDD) for data activities funded by the Millennium Challenge Corporation (MCC), and (ii) facilitate MCC's observance of the general principles of the Federal Policy for the Protection of Human Subjects or "Common Rule [1] ".

For MCC:

- **Transparent** data and documentation refers to the set of practices and tools used to disclose all methods, findings, and data for a data activity. For independent evaluations, this is needed to achieve objectives of evaluator independence, computational reproducibility of analysis, and maximizing broad usability of data for learning.
- **Reproducible** data and documentation refers to facilitating access to code and data required for independent analysts (including MCC staff, staff of the accountable entity designated by a country partner to oversee an MCC-funded program (MCA), and other researchers) to reproduce analysis from shared data with minimal effort. For independent evaluations, this is needed to allow external assessment and understanding of the analysis.
- **Ethical** data and documentation refers to practices that follow the ethical principles of beneficence, respect for persons, and justice, with particular emphasis on informed consent, independent review (institutional or by other review board), and proper data de-identification and management.

Scope

These guidelines apply to all data and documentation produced by independent evaluators contracted by MCC's Department of Policy and Evaluation (DPE) (**MCC staff**) and done in coordination with the country government partners receiving MCC assistance in the form of compacts or threshold program grant agreements (**country partners**). As needed, these guidelines may be referenced for other data activities performed by other **contractors** (evaluators, data collection firms, economic analysis firms, due diligence firms hired by MCC or MCA to conduct MCC-related data collection activities).

TREDD Summary

Staff and contractors need to consider TREDD sharing issues throughout the life cycle of data activities: (i) Design, (ii) Collection and/or Extraction, (iii) Documentation Sharing, (iv) Storage and Transfer, (v) Analysis, and (vi) Data Sharing. Therefore, these guidelines are organized by life cycle stages. These guidelines conclude with a section on managing disclosure risks. A glossary of key terms is provided in Section 11. Although discussed in more detail throughout these guidelines, Table 1 summarizes the TREDD practices MCC requires for each life cycle stage [2] :

Table 1 Summary of TREDD requirements by Life Cycle Stage

Design	Primary data handlers must complete training on protection of human subjects.
	Contractors must have the study documentation (research protocol, informed consent, etc.) reviewed by a US Department of Health and Human Services (HHS)-registered Institutional Review Board (IRB) AND in accordance with any local requirements.
	The informed consent must inform data providers who will have access to what data (identifiable, de-identified) and for what purpose.
	United States (US) and European Union (EU) citizens must be excluded from MCC-funded surveys.
Collection and/or Extraction	Contractors must facilitate and document informed consent with data providers.
	If secondary data owned by another entity is required for analysis, MCC staff and contractors should work toward obtaining Data Sharing Agreements with the data owners to enable public and/or restricted access to the data on the MCC Evaluation Catalog or similar platform.
Documentation Sharing	Contractors must use standard reporting templates unless otherwise agreed with MCC.
	MCC staff must publish relevant documentation of all independent evaluations on the MCC Evaluation Catalog (https://data.mcc.gov/evaluations/index.php/catalog). [3]
Storage and Transfer	MCC staff and contractors must have specific practices in place to protect data confidentiality and data integrity during storage. This includes: encrypting data files; employing password protection on data systems and data encryption; and requiring relevant stakeholders to sign non-disclosure agreements.

	Data handlers must use secure file transfer to transmit digital data files with personally identifiable information (PII).
Analysis	Contractors should establish a reproducible workflow to facilitate computational reproducibility of analysis to the extent feasible. Contractors may consider de-identifying the data prior to analysis to ensure a closer link between the data that produces the analysis and the data that can be shared.
Data Sharing	As feasible and appropriate, Contractors will prepare data and code for public-use and/or restricted-access in a way that adheres to promises of confidentiality made during the informed consent process.
	MCC staff must publish data and code from independent evaluations on the MCC Evaluation Catalog (as feasible).
	Contractors must produce a Transparency Statement in which they confirm (i) public-use and/or restricted-access data and code will reproduce analysis in Interim/Final Results Report, or (ii) justifications for why public-use and/or restricted-access data and code do not reproduce analysis in Interim/Final Results Report.

Background

Accountability, Transparency, and Learning at MCC

MCC is committed to an evidence-based approach for promoting poverty reduction through economic growth. Its results framework seeks to measure and report on the outputs and outcomes of MCC investments. In particular, MCC's Monitoring and Evaluation (M&E) Policy (https://www.mcc.gov/resources/doc/policy-for-monitoring-and-evaluation) is built on the principles of accountability, transparency, and learning:

- **Accountability** refers to MCC's commitment to report on and accept responsibility for the results of MCC-funded activities.
- **Transparency** refers to MCC's commitment to disclose M&E findings in a complete and public manner.
- **Learning** refers to MCC's commitment to improving the understanding of the causal relationships and effects of its interventions, particularly in terms of poverty reduction and economic growth, and to facilitating the integration of M&E findings in the design, implementation, analysis, and measurement of current and future interventions.

In 2013, MCC launched the **Evaluation Catalog** (https://data.mcc.gov/evaluations/index.php/catalog/) to operationalize these principles by creating a platform to transparently share:

- **Documentation** of the independent evaluation portfolio, including the Evaluation Design Reports, Baseline Reports, Interim/Final Results Reports, and Evaluation Briefs, as well as other corresponding documentation (questionnaires, informed consents, data de-identification worksheets, and Transparency Statements). This documentation supports use of the evaluation findings and data by others who may wish to reproduce or extend the analysis of the original evaluation for additional learning.
- **Data** underlying the independent evaluation Interim/Final Results Reports for (i) computational reproducibility and (ii) broader knowledge generation beyond the original evaluation analysis.

While MCC is committed to open data and transparency, MCC has long recognized the need to balance transparency with proper, ethical management of data to minimize risks related to improper data management– in particular data that includes personally identifiable information (PII) and/or sensitive data. The potential risks of improper data management when engaging in data activities may include:

- **Direct harm to data providers from loss of confidentiality**. If intruders or other unauthorized individuals obtain PII or sensitive information that is linkable to the data provider, there is risk that this disclosure could be used to harm and/or exploit the data provider. For example, if the survey is on financial inclusion services and survey participants are identified as loan recipients, with the loan amounts linked to their PII, a loss of confidentiality could result in these individuals – or their households, family members, friends – becoming targets for financial extortion.
- **Reputational harm to data handlers**. Survey firms, independent evaluators, research assistants, and principal investigators all risk loss of reputation if they do not adhere to best practices in ethical data and documentation sharing.
- **Reputational harm to MCC and its country partners**. MCC and the institutions of its country

partners could suffer loss of reputation if data and documentation sharing is considered unethical by other governing bodies, taxpayers, and other relevant stakeholders.

For MCC data activities, [4] these commitments to transparency, reproducibility, and ethical data management creates the need for careful consideration of data management and sharing practices. For this purpose, MCC established the MCC Data Management Guidelines in 2012 to inform proper management of data activities. These TREDD guidelines, effective as of February 21, 2020, supersede and replace all previous versions of the MCC Data Management Guidelines. [5]

Alignment with USG Federal Data Strategy

The mission of the US Government Federal Data Strategy [6] is to fully leverage the value of federal data for mission, service, and the public good by guiding the Federal Government in practicing ethical governance, conscious design, and a learning culture. MCC's TREDD approach reflects the principles of this strategy which include:

Ethical Governance

1. **Uphold Ethics**: Monitor and assess the implications of federal data practices for the public. Design checks and balances to protect and serve the public good.
2. **Exercise Responsibility**: Practice effective data stewardship and governance. Employ sound data security practices, protect individual privacy, maintain promised confidentiality, and ensure appropriate access and use.
3. **Promote Transparency**: Articulate the purposes and uses of federal data to engender public trust. Comprehensively document processes and products to inform data providers and users.

Conscious Design

4. **Ensure Relevance**: Protect the quality and integrity of the data. Validate that data are appropriate, accurate, objective, accessible, useful, understandable, and timely.
5. **Harness Existing Data**: Identify data needs to inform priority research and policy questions; reuse data if possible and acquire additional data if needed.
6. **Anticipate Future Uses**: Create data thoughtfully, considering fitness for use by others; plan for reuse and build in interoperability from the start.
7. **Demonstrate Responsiveness**: Improve data collection, analysis, and dissemination with ongoing input from users and stakeholders. The feedback process is cyclical; establish a baseline, gain support, collaborate, and refine continuously.

Learning Culture

8. **Invest in Learning**: Promote a culture of continuous and collaborative learning with and about data through ongoing investment in data infrastructure and human resources.
9. **Develop Data Leaders**: Cultivate data leadership at all levels of the federal workforce by investing in training and development about the value of data for mission, service, and the public good.
10. **Practice Accountability**: Assign responsibility, audit data practices, document and learn from results, and make needed changes.

Alignment with Scientific Community

MCC's independent evaluations are designed and implemented using research methods from across the social sciences, particularly economics, political science, and other behavioral sciences. MCC's TREDD practices also align with calls for more transparency in the social sciences to mitigate potential threats to the credibility and integrity of the research findings (Miguel et al 2014). **Table 2** provides an overview of the main known threats to credibility and integrity of research and MCC's TREDD practices to mitigate those threats.

An additional threat to the credibility and integrity of MCC-funded independent evaluations is influence – whether actual or perceived – by MCC over the contractors to focus only on positive results of MCC's investments. MCC's TREDD practices discussed in **Table 2** are therefore not only intended to mitigate p-hacking and publication bias driven by researcher and journal practices, but also to protect contractor independence so as to maintain the independence, credibility, and integrity of the evaluation design, implementation, and analysis.

Table 2: MCC's practices to mitigate threats to credibility and independence of independent evaluations 7

Threat	Description	Key references	MCC practice

P-hacking	When analysts, intentionally or not, select a subset of the possible analyses in a study based on whether those analyses generate statistically significant results. The main consequence of p-hacking is that it increases the chances of false positives and can produce biased results within a single study and across a body of literature. The problem can be understood as a version of multiple hypothesis testing where the analyst does not know, or does not report, the true number of underlying hypotheses.	Theoretically outlined in economics by Leamer (1983); Ioannidis (2007) calibrates a model with different levels of p-hacking-type of manipulations by the researchers (among other components) to argue that most published research is probably false; Brodeur et al. (2016) finds evidence of p-hacking in economics using 50,000 tests published in the AER, JPE, and QJE	MCC requires contractors to prepare an Evaluation Design Report. All evaluation questions and corresponding outcomes listed in the EDR must be reported in the Interim/Final Results Report regardless of positive/negative results or statistical significance. Any changes to the evaluation design must be documented and justified in an annex to the original EDR or a new version of the EDR. All reports must follow MCC's reporting requirements. Additionally, all comments made by MCC and other stakeholders on Interim/Final Results Reports, and the response by the contractor, are published alongside the Interim/Final Results Report to mitigate any influence over the contractors to focus on statistically significant, positive findings.

Publication bias	Empirical research suffers from publication bias when results in published studies are systematically unrepresentative of conducted studies. The most common manifestation of such bias occurs when studies with statistically significant results have a higher likelihood of being published than studies with null results.	Franco et al. (2014) found that 22% of studies with null results were published, while 61% of those with strong results were published, in an analysis of studies in economics, political science, sociology, and psychology that were awarded highly competitive resources by National Science Foundation.	MCC requires all independent evaluations to be reported in the MCC Evaluation Catalog as soon as an Evaluation Design Report is cleared. This allows the total number of independent evaluations funded by MCC to be publicly known, even if an evaluation is cancelled. Additionally, all Interim/Final Results Reports are published on the MCC Evaluation Catalog regardless of the reported results and regardless of acceptance into a journal. Summaries of all interim/final evaluations (Evaluation Briefs) are also posted to MCC's main website.

Lack of computational reproducibility	Computational reproducibility is the practice of running the same code over the same data and obtaining the same results as those presented in the original reported analysis.	Gertler et al. (2018) attempted to re-run the analysis code from a sample of 203 empirical papers from leading journals in economics and was able to obtain the same results for 14% of the papers.	MCC requires contractors to submit the analysis code and underlying data. The code and data are published on the MCC Evaluation Catalog. If the public or restricted-use data cannot reproduce analysis (due to data permutations to protect confidentiality for example), the contractor must explain why in the Transparency Statement.

Design

TREDD practices begin in study design, when MCC staff and contractors *define what data needs to be collected for what outputs and outcomes, how, and why.* MCC staff and contractors must determine (i) if the data activity requires collection and handling of PII and/or sensitive data, and (ii) if disclosure of this data may pose any risk of harm to the data provider(s).

If PII does not need to be collected, then it should not be. If data that is being collected is already publicly available and not sensitive, then the necessity of promises of confidentiality should be considered carefully. These decisions should be discussed and agreed between the MCC staff, contractor, and country partners prior to data collection to ensure the research protocol and corresponding informed consent statement(s) align with the requirements of the study.

The following sections describe the TREDD practices to consider during the design phase of the data activity.

Training

The objective of training in the protection of human subjects is for data handlers to understand: (i) key ethical principles in research (beneficence, respect for persons, and justice), (ii) data provider vulnerabilities, and (iii) the risks to data providers, data handlers, MCC, and country partners of improper data sharing, and the corresponding risk mitigation measures. For the data handlers:

- MCC M&E staff, other MCC staff (as applicable) – **Training (with certification from the training provider)** required every 4 years, or sooner in the event of a major change to the Common Rule.
- Contractor Key Personnel – **Training (with certification from the training provider)** required every 4 years [8] , or sooner in the event of a major change to the Common Rule.
- Data Collection/Field Staff – **Training** is required for data collection staff on the informed consent, survey instrument(s), and field protocols established to adhere to ethical principles.
- Other data handlers – **Training** is strongly recommended for MCA staff, other contractor staff who collect, store, analyze, and/or share data.

Understanding Laws and Regulations

Data handlers should identify and understand all relevant local laws for proper data stewardship. Applicable laws include data privacy and protection laws, as well as any national regulations on research and protection of human subjects. There are several resources available to consult and identify relevant laws, including:

- International Compilation of Human Research Standards by HHS is a listing of over 1,000 laws, regulations, and guidelines on human subjects' protections in 130 countries and from many international organizations.
- Data Protection Laws of the World by DLA Piper Law Group and Data Protection around the World by Commission Nationale de l'Informatique et des Libertés (CNIL) allow users to compare laws and regulations between countries.

The applicable laws and regulations may evolve over time. MCC and MCA Office of General Counsel (OGC) staff can support data handlers in understanding these issues as needed.

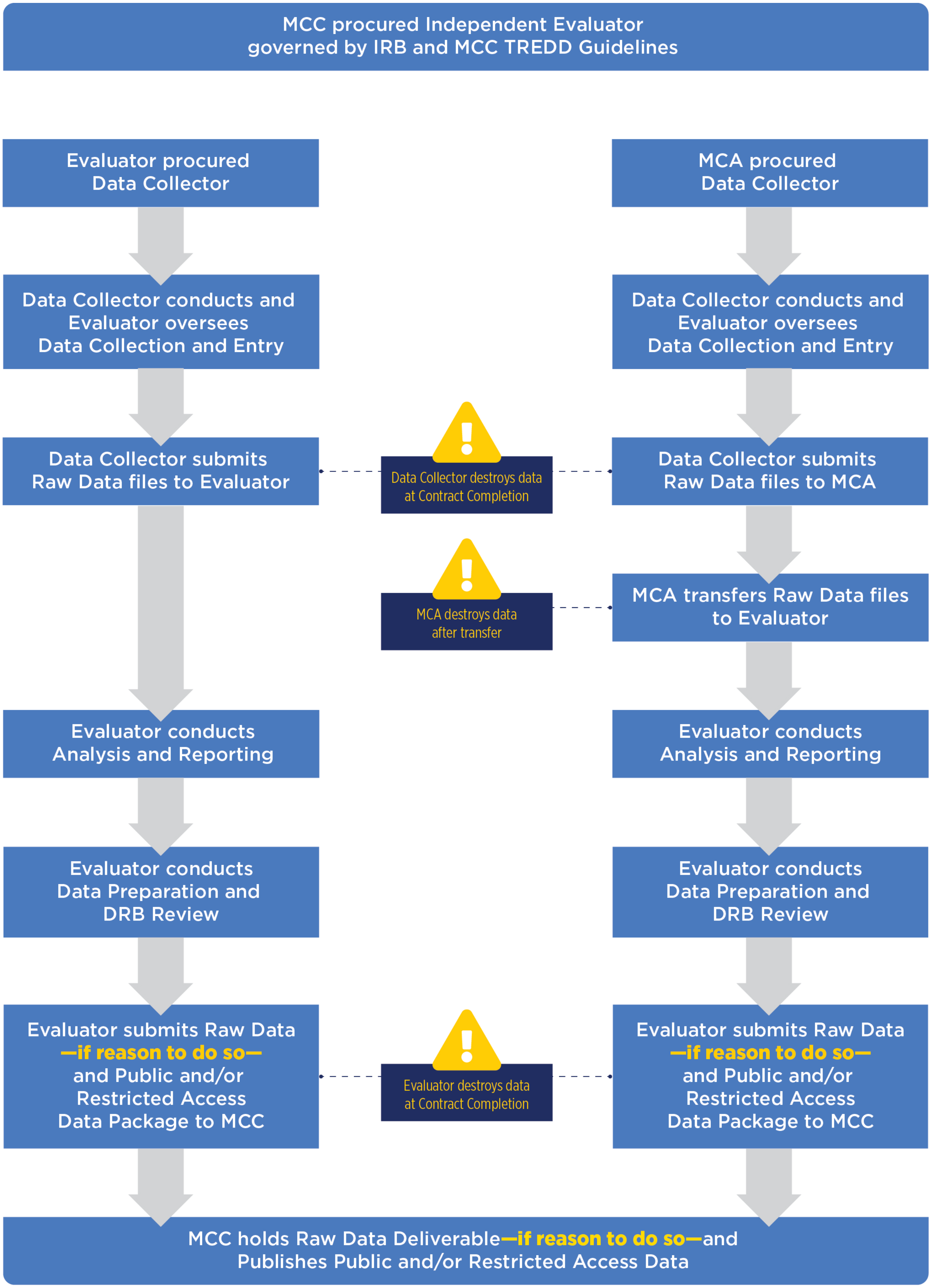
Contractors are also required to exclude United States (US) and European Union (EU) citizens from MCC-funded surveys. This is to mitigate additional requirements for managing data in accordance with US and EU data privacy laws.

Identifiable Data – Handlers and Data Flow

Early in the Design stage, MCC, country partners, and contractors should identify the necessary data handlers over the course of the data activity life cycle and clearly document who needs access to identifiable data and when. This information should be determined prior to submission of the research protocol to the Institutional Review Board (Section 3.4). Depending on the data activity and/or procurement mechanisms, there can be the multiple data handlers, including MCC, MCA, evaluation firms, and data collection firms.

Figure 1 outlines the two common data workflows across data handlers for an independent evaluation to consider when building the research protocol and informed consent. However, MCC notes that data handlers and data flow are context specific and may be adapted to the needs of the study and the requirements of the Institutional Review Board.

Figure 1.



MCC procured Independent Evaluator governed by IRB and MCC TREDD Guidelines

Evaluator procured Data Collector

Data Collector conducts and Evaluator oversees Data Collection and Entry

Data Collector submits Raw Data files to Evaluator

Evaluator conducts Analysis and Reporting

Evaluator conducts Data Preparation and DRB Review

Evaluator submits Raw Data —if reason to do so— and Public and/or Restricted Access Data Package to MCC

MCA procured Data Collector

Data Collector conducts and Evaluator oversees Data Collection and Entry

Data Collector submits Raw Data files to MCA

MCA transfers Raw Data files to Evaluator

Evaluator conducts Analysis and Reporting

Evaluator conducts Data Preparation and DRB Review

Evaluator submits Raw Data —if reason to do so— and Public and/or Restricted Access Data Package to MCC

Data Collector destroys data at Contract Completion

MCA destroys data after transfer

Evaluator destroys data at Contract Completion

MCC holds Raw Data Deliverable—if reason to do so—and Publishes Public and/or Restricted Access Data

2020-017-2343-01

While it is not MCC's practice to routinely hold raw, identifiable data after the completion of a data activity, it may do so in a limited number of circumstances. For example, MCC may hold identifiable data: (i) to facilitate the transfer of an evaluation task from one contractor to another (where contractor 1 submits raw data to MCC to transfer to contractor 2), (ii) if there is a specific and known need to revisit same data providers for future data activities (such as to study sustainability issues), and (iii) if other business requirements are identified by the DRB (see Section 9.5). For this reason, contractors should ensure informed consent and research protocols allow and facilitate MCC's holding of identifiable data, or if that is not feasible, establish protocols for how the contractor will manage transfers to another contractor or other MCC designated agent should the need arise. Where MCC does have reason to hold identifiable data, contractors are requested to submit the identifiable data to MCC only (not for access outside of MCC). In addition, as the holder of identifiable data, MCC will continue to protect the confidentiality of such data and withhold data where the disclosure of such is prohibited by law or MCC reasonably determines that the disclosure of such would harm an interest protected by an exemption under the Freedom of Information Act (FOIA) [9] .

Institutional Review Board (IRB)

According to US regulations, IRBs "*assure, both in advance and by periodic review, which appropriate steps are taken to protect the rights and welfare of humans participating as subjects in the research. To accomplish this purpose, IRBs use a group process to review research protocols and related materials (e.g., informed consent documents and investigator brochures) to ensure protection of the rights and welfare of human subjects of research.* [10] "

All MCC-funded independent evaluations that require human subjects – whether quantitative, qualitative or both – are required to undergo IRB review, even if they are exempt under HHS definitions of human subjects' research. This requirement is built into the Standard Evaluation Firm Scope of Work. **If an IRB initially classifies the evaluation as "exempt" from full review, the contractor must discuss with the MCC Project Manager (PM) and Contracting Officers Representative (COR) how to proceed given MCC's preference for all independent evaluation protocols to be reviewed by an IRB**. If the contractor believes the independent evaluation should not undergo IRB review, a justification must be submitted to MCC and must be cleared by the MCC PM, COR, and M&E Managing Director.

For MCC data activities, there are three main types of IRBs to consider:

- National IRB – This is a centralized IRB established within a country to review and govern research in that country.
- Academic Institution IRBs – IRBs that are based within universities to govern the research produced by university staff. This is typically required if one or more staff members of the contractor are based in an academic institution.
- Independent IRB firms – There are independent IRBs that may be contracted for academic and non-academic research.

Working with MCC staff and MCA staff, contractors will lead the IRB process for relevant data activities, which consists of the following steps:

- **Identify IRB(s)** –MCC requires its contractors to submit to an HHS-registered IRB AND adhere to any National IRB requirements (as applicable). Depending on the context, the contractor may need to submit a protocol to multiple IRBs (for example, if a National IRB is required, but it is not HHS-registered, the contractor will also need to submit to an HHS-registered IRB). Staff and contractors can reference the International Compilation of Human Research Standards and Office for Human Research Protections (OHRP) Database to identify appropriate contacts and IRBs to ensure the protocol is reviewed by at least one IRB that is HHS-registered and local requirements are followed.
- **Timing** – The contractor should identify in advance the schedule(s) for the IRB(s) and the time requirements for submission and review. These may vary significantly across contexts and type of IRB. Given this process can take 1-3 months or more, depending on the IRB process, contractors should build this into their projected timeline as early as possible.
- **Cost** – The costs of initial and periodic IRB reviews vary by country and by how many years the protocol must be in place. This is because standard IRB review may require both an initial, larger submission and review fee, as well as an annual review fee to maintain the IRB coverage over the course of the research life cycle. For reference, costs for initial reviews by Independent IRB firms can range from $1300-$1500 USD, with annual review fees between $800-$1500 USD. This can be reduced if the study is determined to be exempt or not required to fall under annual review.
- **Representation** – Depending on IRB requirements, contractors may need to submit and/or present the research protocol and documents to the IRB in person. This should be built into the work plan and budget accordingly.

Preliminary Findings

Early in the design stage when preparing the IRB review package, contractors should determine whether a timely feedback loop is required to share data activity findings. MCC has identified cases in which it was critical to share results of the data activity (even preliminary results) with data providers and/or other stakeholders in a timely way because the data collected directly affected the health and well-being of the population. MCC and contractors should assess upfront what data is being collected and, if applicable, whether data that directly affects the health and well-being of the population can be reported back in a timely manner, and through what mechanism(s). MCC views such ethical responsibilities as superseding any methodological concerns about contaminating the study sample. When such data is being collected – for example water quality testing – a process for ensuring an appropriate feedback loop should be built between MCC, the contractor, and relevant local stakeholders before the results are available. **This issue should be fully discussed in the research protocol and agreed with the IRB before proceeding to data collection.** In addition, this information sharing may require additional financial resources in the evaluation budget.

Informed Consent

The informed consent should be context-specific and furnish data providers with the following information:

1. Statement on data activity purpose (i.e. program evaluation) and voluntary nature of their participation.
2. Duration and description of specific procedures, reasonable expected risks of providing data, and

reasonable expected benefits of providing data.

3. Promises of confidentiality and data sharing. The contractor should first determine if (i) PII data needs to be collected (for specific study purposes) AND (ii) if confidentiality promises are required. If PII data is not needed – it should not be collected, thereby limiting specific risks to confidentiality. If the data collected is public – directly observable and not sensitive – promises of confidentiality should be carefully considered as they may be unnecessary. The following statements in the informed consent lay the (possible) foundation for proper data sharing in the future:
    1. Statement of whether or not the data will be shared, and if shared, with whom and to what extent. In particular, it may be necessary to clarify who will have access to the identifiable dataset and who will have access to a de-identified version of the dataset. If computational reproducibility requires access to identifiable data, then a statement on who will have access to identifiable data for the purpose of reproducibility should be included. If de-identified data will be made public or otherwise shared, then the statement should state this [11] ;
    2. Statement on how data will be de-identified (as applicable); and
    3. Broad consent – The contractor may consider obtaining broad consent for identifiable data to be shared with other researchers for unknown learning purposes [12] . However, even if broad consent is obtained, MCC anticipates minimizing holding and/or sharing identifiable data to mitigate unauthorized disclosures or misuse of identifiable data that may cause harm to data providers and/or handlers.

MCC provides contractors with a generic informed consent statement template (**Annex 1**). The MCC PM and contractor should review the draft informed consent language (preferably based on the template) to ensure agreement prior to submission to the IRB. However, MCC recognizes the final informed consent statement **must be reviewed and cleared by the contractor's IRB(s)**. In instances where the IRB requires changes to the informed consent which may limit MCC's ability to collect, store, and/or disseminate relevant data, contractors must notify MCC of these changes by providing a copy of the IRB approved informed consent marked to show changes from the originally agreed language. MCC staff will determine whether the IRB's required changes should be discussed with the MCC Disclosure Review Board (DRB).

Future Data Sharing

Data handlers should carefully consider how elements of the design may inform (or prevent) future efforts for data de-identification and data sharing and take appropriate actions, including but not limited to:

- **Monitor knowledge of sample frame**. What is the source for the sample frame, how available is this source to others, what is the size of the sample frame, and what percentage of the sample frame will be selected for the study? These are questions that will inform data de-identification efforts, specifically focused on understanding the extent to which outliers in the study sample may be outliers in the population, and therefore potentially useful for re-identification of individuals, households, communities, etc.
- **Monitor availability of linkage documentation.** For future de-identification efforts, the data handlers will need to be aware of linkage documentation that may support re-identification efforts or mitigate de-identification efforts. For example, when preparing a public-use data set, the contractor may determine village names need to be de-identified and removed from the data. But if the names of the villages in the sample are disseminated elsewhere, such as in an Evaluation Design

Report, this information could be used to re-identify the village names in the dataset and increase household/individual re-identification risk. For this reason, MCC staff and contractors should carefully consider what information is and will be available about the study sample that may pose a re-identification risk and manage appropriately.

- **Monitor knowledge of treatment**. For program evaluations, how well-known is "treatment" status? For example, will random selection of communities/villages/schools/facilities/etc. receiving the treatment be publicized? This should be carefully considered as it is a form of linkage that may support re-identification efforts if the treatment status, or other information about the treatment group, is known and can help to re-identify data providers. MCC staff and contractors should therefore consider carefully how treatment status, program beneficiary lists, etc., may be managed to mitigate future re-identification risk.
- **Consider de-identification strategy early**. De-identification efforts often require data permutations – such as suppression of specific variables' values, including top and bottom coding, conversion of continuous variables to categorical or removal of any identifiable variation. Even if data does not need to be submitted to MCC until all data rounds are completed, MCC requires contractors to begin documenting their de-identification strategy in the De-Identification Worksheet (discussed below) at the completion of each round of data collection, as per the Standard SOW.
- **Flag identifying and sensitive data.** Beginning with questionnaire design and data entry, the contractor may consider creating flags – such as a specific suffix in the variable number or name – to create an easy reference in data analysis, de-identification, and dissemination for variables which should be carefully considered. These variables may then be removed or permutated for proper data sharing in adherence with promises of confidentiality and risk mitigation.

Documentation Sharing

For independent evaluations, as soon as an Evaluation Design Report is cleared by MCC staff, MCC staff and the contractor will begin documentation sharing. As per the Standard SOW Section F.3 Deliverables, a sub-set of study documentation will be posted on the MCC Evaluation Catalog and must be **Section 508** [13] compliant (https://www.section508.gov/). **Table 3** summarizes the required documentation and format for study documentation that must be made publicly available.

Table 3: Documentation Required for Sharing

Document	**Requested Format**	**Description**
*Metadata File***(Annex 2)**	Nesstar file;PDF for viewing purposes	The contractor should prepare the metadata file for the public evaluation catalog entry. The metadata can be updated/revised as necessary over the course of the evaluation. Contractors should not attach any data sets or related documents under the “other materials” or “external resources” sections. Data is reviewed, cleared, and posted separately as per the DRB review process detailed in later sections.Please note, MCC reviews the PDF export of the Nesstar file and recommends contractors review this PDF export prior to submitting to MCC.
Evaluation Design Report, Baseline Report (as applicable), Interim Report(s) (as applicable), Final Report, any relevant presentation materials	Word or searchable PDF	These documents (deliverables required under MCC contracts) provide necessary design and analytical information. Contractors should ensure that all public use documents/reports have been reviewed and edited to remove any references,

Document	Requested Format	Description
Cost-Benefit Analysis (CBA) Model	Excel or other relevant format	such as geographic locations, that may threaten or undo data de-identification efforts. MCC requires contractors to update Evaluation Design Reports (EDRs) as needed over the life of the evaluation. Any revisions should be documented in the EDR so that course corrections/revisions are clearly documented. In the event that one contractor inherits an evaluation from another, the original contractor's EDR will be posted on the Evaluation Catalog along with the new contractor's EDR.
Informed Consent Statement	Word, searchable PDF	The IRB approved informed consent statement should be published, either independently or as part of the questionnaire(s).
Questionnaires (English and local language) and related documentation	Original editable source and searchable PDF	All survey questionnaires – baseline, interim, final – should be shared as the original editable source file. Contractors may also submit a searchable PDF. Related documentation may also include sampling strategy, field operations and interviewer manuals when needed for complete documentation of survey protocols. Any translation requirements should follow the contractor scope of work. For qualitative data, this documentation should include the interview guide(s) and any other study materials necessary

Document	Requested Format	Description
		for understanding how the data was generated and analyzed (as feasible).

Contractors may be required to submit an 'internal only' version of a document, as well as a 'public-use' version of the document in order to mitigate the public release of linkage documentation that could support future re-identification of the publicly available data. For example, if the Design Report contains geographic identifiers that may enable future re-identification of the data provider(s), that information may be included in an internal-only version but must be removed from the public document.

Registries

In addition to the requirement to publish documentation and data on the **MCC Evaluation Catalog**, contractors may also choose to register the data activity on other study registries:

Table 4: Optional study registries

Registry	Notes
AEA Registry (socialscienceregistry.org)	Registry for randomized control trials (RCTs) in economics
Clinical Trials or ICTRP (clinicaltrial.gov)	Registry for randomized control trials (RCTs) in health-related fields
3ie (ridie.org)	Registry for randomized control trials (RCTs) and quasi-experimental designs in development economics/program evaluation
EGAP (egap.org)	Registry for experiments and observational studies in governance and politics
OSF (osf.io)	Registry with multiple formats: short, long, structured, and open ended for any method, across social sciences

As feasible, any additional registries for independent evaluations should be directed back to the **MCC Evaluation Catalog** as the centralized source for all documentation and data associated with the evaluation.

Contractor Reporting Guidelines

Reporting guidelines are a standardized procedure to report on study design, implementation, analysis, and interpretation of findings. For MCC, the evidence generated by any single independent evaluation is intended to contribute to a body of knowledge – such as MCC's Principles into Practice (https://www.mcc.gov/our-impact/principles-into-practice) papers, systematic reviews, and other knowledge products. To facilitate this, MCC provides contractors with templates for reporting requirements for Design, Baseline, and Interim/Final Reports (**Annex 3**). Any deviations from the standard report templates should be discussed and agreed between MCC staff and the contractor prior to developing the report. The goal is for MCC-funded evaluations – and other data activities – to be accessible for broader learning and systematic reviews, and to serve as inputs to future cost-benefit analysis models.

Data Collection and/or Extraction

Data Use, Transfer, and Sharing Agreement(s)

In the design stage, contractors should identify all required data sources for the study. As part of this process, contractors may identify existing data sources that may provide cost-effective input into the data activity. In these cases, contractors should work with MCC staff – including MCC's legal counsel and country partners – to develop a documented Data Sharing Agreement between the contractor and the owner of the existing data source. The agreement should include documented understanding of (i) who owns the data, (ii) whether the contractor or data owner, can prepare the data for public and/or restricted access use, and (iii) whether the data can be made available through the MCC Evaluation Catalog or other mechanisms (such as the country government data platforms).

If existing data is extracted/obtained by a contractor for analysis, but an agreement with the data owner was not or cannot be put in place to facilitate future preparation and access to that data, the contractor must document this in the Transparency Statement (See Section 8.4).

Remote-sensing Imagery

During the Data Collection phase, contractors may also need to access imagery or other remote sensing data. Such data may, for example, be used to estimate outputs or outcome variables of interest such as cropped area or yields, or the extent and nature of built infrastructure. Remote-sensing imagery includes at least two types of data: satellite and drone imagery.

If contractors believe satellite imagery is required for their analysis, they are directed to **Annex 4,** Satellite Imagery Data Requests and Management, which details how contractors may use using existing MCC mechanisms to obtain satellite imagery. In addition to this mechanism, additional guidance on access and use of other satellite and drone imagery may become available in future versions of these guidelines.

As with other forms of data, if the satellite and/or drone imagery used in the analysis entails limited or restricted access due to privacy or ownership issues, the contractor should detail this in the Transparency Statement (See Section 8.4).

Data Storage and Transfer

Paper-based Data Storage and Disposal

When used, paper-based questionnaires are often saved during the data entry phase and initial data analysis phase for verification of survey responses. However, such verification should be addressed as soon as possible alongside data entry. Double-data entry, with paper-based verification of discrepant responses, yields nearly perfect correspondence between responses and keypunched data and should be required of all *quantitative* surveys. Once data entry and verification of survey responses has been completed, all the paper-based questionnaires should be immediately and securely destroyed (shredded or burned depending on local resources). When applicable, contractors should determine if other documents (informed consents, other documentation verifying how study protocols were implemented, etc.) should be scanned and stored digitally before destroying paper-based versions. Disposal notification is included as a final deliverable of the contract.

Digital Data Storage and Disposal

Once data are entered, there should be specific practices in place to protect data confidentiality and integrity while the data is stored digitally, such as: encrypting data files; employing password protection on data systems and data encryption; and requiring relevant stakeholders to sign non-disclosure agreements. As per MCC information technology standards, the **end point encryption software should meet AES-256 encryption standards or above**.

As discussed in Section 3.3, once a data handler's role is completed, the data handler is required to ensure appropriate disposal of the digital data. To prevent unintentional release, the contractor must provide media sanitization procedures for the clearing or purging of all media that holds or has held relevant PII data in accordance with *NIST SP 800-88, Guidelines for Media Sanitization* [14] . Overwriting media by a US Government approved technology, method, or tool is acceptable. Sanitization procedures will need to be approved by MCC. The contractor must provide written attestation to MCC by contract closure of the media sanitization for any PII data generated.

Digital Data Transfer

When sharing data files, data handlers should use a secure file transfer (SFTP) system and should control access to the storage mechanism. The following techniques should be considered:

- Encrypt all communication channels, especially over Wi-Fi connections;
- Limit Wi-Fi connections to trusted parties; avoid public locations, if possible;
- File transfers should occur only through HTTPS connections;
- Use of hyperlinks for connections should be prohibited; instead, users should only connect to trusted sites by manually starting a new web-browsing session; and
- As a last resort, password protect and encrypt all PDFs or other document types if there are no other solutions available for secure file transfers. Send passwords via a separate email or phone the recipient.

Analysis

Considerations for Reproducibility

Depending on the requirements of the data activity, contractors may conduct analysis at many points in time – baseline, interim, and/or final. Contractors should consider the following:

1. **When there are multiple rounds of data (baseline, interim, final), MCC prefers all data to be prepared in one complete data package for data sharing.** MCC aims for public and/or restricted-access data to be as complete as possible. This means all data that was collected as part of the data activity is included in the data package (not just constructed variables produced for the analysis report or just sub-sections of questionnaires used in final analysis). Unless otherwise agreed with MCC staff, contractors should plan to package all data collected from all data rounds (baseline, interim(s), and final) as one data package. This is to ensure consistency in how de-identification of data is managed across data rounds, minimize risk of re-identification across rounds, and reduce costs.
2. **Establish reproducible workflow.** In accordance with the contractual requirements, contractors should establish and maintain a reproducible workflow for analysis to ensure a direct link (as feasible) between the future public and/or restricted-access data, the analysis code, and the analysis results presented in baseline, interim, and/or final analysis reports. [15]
3. **Separate de-identification code from analysis code**. As a standard contract deliverable, MCC requests analysis code (code written in statistical software program to produce analysis) submitted as part of the final data package. This means the contractor should ensure any de-identification code is written separately from analysis code because de-identification code will not be publicly shared.
4. **Run analysis code on de-identified data**. When possible, contractors should run analysis code on the de-identified data files to demonstrate reproducibility successes and/or challenges. This would improve documentation associated with reports and data, and inform the Transparency Statement to report what can, and cannot, be reproduced using the public-use and/or restricted-access data**.**

The Standard Evaluation Firm SOW provides specific detail on contractors' requirements for ensuring appropriate review, feedback, and dissemination of analysis reports.

Data Sharing (Preparation)

After the Final Report for an evaluation is published, per the terms of MCC's Standard Evaluator SOW, contractors are allowed a period of exclusivity not exceeding six (6) months during which only they will have access to the data package. This exclusivity period facilitates contractors' completion of academic articles and other analysis prior to allowing new researchers access to the data. However, MCC aims for the full data package (all rounds of data) to be accessible **no later than 6 months** following publication of the **Final Report; and** any extension of the exclusivity period beyond six months requires approval from the MCC PM and COR. In any event, contractors should complete data preparation, review, and clearance before the end of the contract period of performance.

Data Package

MCC anticipates that data from each evaluation – or similar data activity – may fall into one of the following categories:

- **Public-use data**. This is data that has been **de-identified** or **does not require de-identification** and may be shared publicly without posing a risk of harm to the data providers. For independent evaluations, this data will be available for direct download from the MCC Evaluation Catalog.
- **Restricted-access data**. This is data that may contain identifiers (direct and/or indirect) requiring that any sharing of the data be subject to conditions that MCC determines in its discretion are appropriate to protect the data provider's confidentiality. MCC currently does not share data on a restricted-access basis.
- **No access data**. This is data that cannot be sufficiently de-identified so as to be made accessible through either public-use or restricted-access. When preparing a No-Access data file, the Contractor and PM should work together to determine what Data Package Requirements should be submitted. In some cases, a full Data Package may still be submitted for full documentation of the decision to have No-Access, for some cases, it may be sufficient to notify MCC it is No-Access data and provide the Transparency Statement.

When ready to prepare data for public-use and/or restricted-access **contractors should expect to prepare and submit the following package to MCC**:

Table 5: MCC Data Package Requirements

Element	Requested Format	Description
DRB Data Package Worksheet	Word**(Annex 5)**	Contractors will complete this worksheet to document the actions taken to de-identify and prepare the data for public and/or restricted-access use.
Data – Clearly labeled as (i) Public Use, (ii) Restricted Access,	Stata 13 *(or other format agreed with MCC)*	This should be the complete data file(s) –

Element	Requested Format	Description
and/or (iii) No Access (if justified)		including the full dataset as collected (required) and any constructed analysis variables (optional – it is assumed analysis code will produce these). The ability to de-identify the data as per informed consent promises will inform whether or not this data is public use, restricted access, or no access.
Data Codebook– Public Use and/or Restricted Access only	PDF	Stata codebook output to review data – the codebook should include a label book as well as basic summary statistics including frequency and distribution information.
Analysis Code	Stata do file	This is the analysis code to produce the variables and analysis reported in the analysis report(s).
Transparency Statement	Searchable PDF	Contractors should prepare a Transparency Statement which states the extent to which data (public use and/or restricted access) can enable computational reproducibility of results presented in report(s).

If necessary, this package should also include any updates to the Metadata for the MCC Evaluation Catalog.

Data De-Identification

To adhere to promises of confidentiality made during the informed consent process and to mitigate risks to data providers for providing PII and/or sensitive data in the data package, data that is prepared for public-use must be de-identified. For restricted-access use data, the level of data de-identification may vary depending on promises of confidentiality made. Prior to conducting data de-identification actions, contractors should:

- Consider risk factors and probability of re-identification as presented below in **Table 6**.
- Maintain a balance between applying data perturbation-based methods and techniques to de-identify data and ensuring the quality, usability, and relevance of the data. In many cases, significant de-identification efforts may result in data that is less useful and/or relevant, even for computational reproducibility of original study analysis.
- Carefully consider combinations of variables, even when individual variables do not pose a re-identification risk. For example, age, gender, or marital status alone may not pose re-identification risk, but when combined these variables may be sufficient to identify the data provider, resulting in a re-identification risk.

Table 6 –Re-identification Risk Factors and Probabilities

Risk Factor for re-identification	Lower probability	Higher probability
Sample representation: Are outliers in the data outliers in the general population?	When the sample is a small percentage of the general population, visible and known characteristics that are outliers in the sample may not pose a re-identification risk because there are other similar individuals/households/businesses/etc. in the sample frame	When the sample is a large percentage of the general population, visible and known characteristics that are outliers in the sample may pose a stronger re-identification risk because there are few to no other similar individuals/ households/businesses/etc. in the sample frame
Linkage documentation: What documentation about the sample exists outside the research data but can link to it?	If little to no documentation exists about the study sample, then linkage documentation may not pose a re-identification risk	If documentation exists about the study sample, then linkage documentation may pose a re-identification risk (examples: loan information obtained on study sample mirrors loan information at bank)
Timing and population characteristics: How closely does the data reflect current and future state for the sample population?	If significant time has passed and the study population is transient or nomadic, there is lower re-identification risk	If the data was recently collected and the study population is more permanent, there is higher re-identification risk

Once the above has been considered, contractors may consider the following high-level data perturbation techniques for data de-identification:

- **Removal of all direct identifiers.** Removal of direct identifiers may not be as simple as removing the specific variables where known direct identifiers were recorded by the survey team. For example, the written response within “Other” responses may include direct identifiers.

- **Geographic units**. Contractors should consider the highest geographic level that should remain identifiable for specific analytic purposes and de-identify all lower geographic units. Similar to the discussion above on sample representation, the higher the geographic unit that is de-identified, the lower the risk for re-identification at individual, household, and other sample unit levels and often less data permutation is necessary on a variable-by-variable basis.
- **Top and Bottom Coding**. When specific continuous variables are visible and/or known characteristics about the data provider (i.e. visible asset holdings, age, years of education), outliers may need to be considered for top and bottom coding. There is no specific rule (top and/or bottom 2%, 5%, etc.) given the decision on where to cut outliers should be made based on the data and what is known about the study sample population. To retain data values and avoid lost data, contractors can send outlier values to the median once a threshold is identified.
- **Re-categorization**. When specific categorical variables are visible and/or known characteristics about the data provider (i.e. ethnicity, religion, language spoken, education level), minority groups may need to be considered for re-categorization. To retain the value of the data, it's preferable to re-categorize into meaningful groups, combining categories, rather than collapsing into an unknown "Other" category. However, this is dependent on context, data, and risk.
- **Removal of indirect identifiers**. When specific variables cannot be retained given potential re-identification risk, the variable(s) should be removed from public-use datasets (and clearly documented as removed).

Qualitative Data

As of February 2020, typically MCC does not expect qualitative data to be prepared for public-use given unknowns [16] regarding de-identification and usability of qualitative data. However, if contractors determine it is feasible and appropriate to prepare the data for restricted-access, they should work with MCC on determining whether and how to proceed with data preparation for sharing.

All relevant documentation should still be shared for public dissemination as per **Table 3**.

Transparency Statement

Contractors should prepare a Transparency Statement which states the extent to which data (public use and/or restricted access) can or cannot reproduce the results presented in the evaluation report. This will be discussed with the DRB and then finalized based on the final approved data file(s). Contractors may reference **Annex 6** as a template.

Data Sharing (Process)

Disclosure Review Board (DRB)

The MCC Disclosure Review Board was established in 2013 with three primary responsibilities (**Annex 7**):

1. To develop, review and approve guidelines and procedures (including modifications thereto) for data activities;
2. To review and approve proposals related to data disclosure; and
3. To notify the MCC Incident Response team in the event of an identified, specific disclosure risk (spill, breach, etc.) and follow MCC protocol for risk management.

DRB Submission and Review Process

The contractor responsible for the data activity will prepare the Data Package for DRB review, which involves a multi-step process (**Annex 8**):

1. **Set Date** – Contractor and M&E Project Manager (M&E PM) agree on expected DRB review date as early as possible to confirm scheduling in line with the contract and work plan. Given the DRB review process, this should be scheduled at least 1-2 months before the contract expires.
2. **PM Review** – Contractor should submit the full Data Package to the M&E PM. The M&E PM should review the Metadata, Data Package Worksheet, and Transparency Statement for clarity and completeness. This may require one or more rounds of revision based on the M&E PM requests for clarity and completeness.
3. **M&E Technical Review** – The M&E PM and the M&E DRB members will conduct a technical review and provide feedback to the contractor on the proposed data de-identification process. This may require a second round of revision to the package based on feedback on documentation clarity and completeness, as well as the proposed de-identification strategy.
4. **Submission for DRB review** – Following revisions based on the technical review, contractors should re-submit the full Data Package (including public-use and/or restricted-access data) to the M&E PM for submission to the DRB at least 2 weeks prior to the agreed DRB review date.

For a DRB review, the contractor will present an overview of the study, the proposed data de-identification approach and other necessary activities for data sharing, and respond to any questions from the DRB. **A decision whether or not to share the data may be made during the meeting or may require additional follow up by the contractor. Depending on the context and risks, the DRB may determine it is not possible to de-identify a dataset sufficiently to allow for public and/or restricted-access use.** All DRB decisions are documented in the DRB minutes.

If any feedback/revisions are required following DRB review, the contractor will revise and resubmit the full data package to the M&E PM with documented responses to DRB feedback to ensure timely, updated review and clearance of the full package prior to public (or other) posting.

Public-use Data Sharing

Once cleared by the DRB, any public data will be immediately posted, and available for direct download from the MCC Evaluation Catalog.

Restricted-access Data Sharing

MCC currently does not have a restricted-access data sharing mechanism. All data prepared for restricted-access use will be held by MCC until a data sharing mechanism is established. This data will then be reviewed again in accordance with an established restricted-access data sharing protocol.

No access Data

The DRB may determine that the data de-identification efforts of the contractor are insufficient to adhere to promises of confidentiality and therefore the de-identified data cannot be made accessible through either public-use or restricted-access. In such cases, the Transparency Statement will be updated to reflect the DRB decision.

"No Access" data that has already been submitted to MCC will be treated as Identifiable Data as per Section 3.3.

Identifiable Data Sharing

MCC's data sharing practice primarily involves the sharing of public-use data. However, there are cases in which MCC may facilitate access to identifiable data, particularly if there is a critical business need for access to the data before the data can be prepared for public-use (such as an input into a Cost-Benefit Analysis Model) and the sharing of such data conforms with the applicable informed consent. Any request to share identifiable data must be submitted to the DRB using **Annex 9** Identifiable Data Sharing Form. The DRB may, in its discretion, permit access to such data after reviewing the request.

Unauthorized Disclosure Management

Types of Disclosure Risks

Unauthorized disclosure may occur during data collection and storage (through lost or stolen computers, USB drives, computer hacking) or through dissemination of public and/or restricted access data. There are several types of risks related to unauthorized disclosure of data that contains PII and/or sensitive data:

- **Low risk:** Disclosed data may be linkable to other data and/or documentation that could serve to support re-identification of individuals, households, firms, etc.
- **Medium risk**: Disclosed data includes indirect identifiers that could support re-identification of individuals, households, firms, etc.
- **High risk:** Disclosed data includes direct identifiers that will identify individuals, households, firms, etc. In the event of a high-risk disclosure, the contractor should anticipate conducting a full risk assessment of the data disclosed.

Risk Mitigation and Management Process

In the event of an unauthorized disclosure of data, the following steps must be taken in addition to any reporting by the contractor required under its IRB protocol [17] :

1. If it is the contractor who identifies the unauthorized disclosure, their representative must notify their respective **MCC PM** The **MCC PM** must notify the **DRB** of any disclosure incident immediately.

2. If applicable, MCC will immediately remove the respective dataset(s) from the MCC Evaluation Catalog.

3. **The contractor,** working with the **MCC PM,** will have one week starting from the notification to the DRB to complete the Disclosure Incident Form (**Annex 10**). Depending on the nature of the disclosure, this may include a full risk assessment of all data disclosed, as well as a revised Data Package.

4. The DRB will convene to review the disclosure incident documentation. Standard procedure will be:
   1. DRB Chair will notify the Incident Response Team;
   2. DRB M&E members will request an independent risk assessment by relevant MCC DCO, M&E, and MCA staff – This will require country and sector specific knowledge.
   3. DRB M&E members will request an independent quality assurance of data package preparation consisting of a review of the de-identification process and assessment of remaining risk prior to re-submission to the DRB.

5. For any data disclosures, the DRB will work with appropriate stakeholders to determine whom to notify, both internally within MCC and with respect to country partners.

6. The DRB will convene to review the final, complete disclosure incident documentation package,

including the independent risk assessment and independent quality assurance of data package preparation. Decisions on how to proceed will be made with the Incident Response Team and recorded in the DRB Minutes.

Glossary

beneficence: an ethical principle of research that incorporates two ideas: (i) do no harm and (ii) maximize possible benefits.

computational reproducibility: the practice of running the same code over the same data and obtaining the same results as those presented in the originally reported analysis.

contractor: any firm or individual hired by MCC or an MCA to conduct a data activity.

country partner: as defined in Section 1.2, each country government partner receiving MCC assistance in the form of a compact or threshold program grant agreement.

data: individual, household, community, contextual, and entity-level information that MCC and its country partners collect, produce and/or use to inform investment decisions, operations, or monitoring and evaluation activities for MCC-funded assistance programs.

data activity: any action involving the designing, collecting, storing, analyzing, or sharing of data (e.g., the conduct of an independent evaluation is a data activity).

data confidentiality: Measures taken to maintain data confidentiality including, but not limited to, data encryption; maintaining an authorized access list and/ or requiring non-disclosure agreement(s); and knowing and possessing authorized rules for handling, storing, and transferring data with approved methods (e.g., encryption).

data de-identification: general term for any process of removing the association between a set of identifying data (direct and indirect identifiers) and the data provider. De-identification includes all techniques that allow access to data while simultaneously limiting the opportunity for unwanted disclosure.

data handler: any person (individual or legal entity) who collects, stores, analyzes, and/or shares data.

data integrity: the accuracy and consistency of the data, ensuring the data is unchanged, intact, and complete. Data integrity is achieved by protecting data confidentiality, authenticity, and limiting modification to authorized users or events.

data perturbation: methods used to alter data in order to mitigate risks to data provider (i.e. removal of PII/sensitive data; top/bottom coding of outliers)

data provider: any individual, household, community, or other entity who provides data.

direct identifiers: data that directly identify a person (individual or legal). This data may include full name, date of birth, mailing or home address, email address, telephone number, GPS coordinates, national identification number, and physical/biological identifiers (e.g., physical appearance, through photo or video data collection, fingerprints, DNA, etc.). Depending on the study and data needs, direct identifiers can also include the name of the school, health facility, community, etc. that directly identify the location of the data collection or extraction.

documentation: written materials that disclose the methods behind data activities, including but not limited to Design Reports, Baseline Reports, Interim/Final Reports, Evaluation Briefs, questionnaires, Transparency Statements, Statements of Difference/Support, peer review comments and responses.

Disclosure Review Board (DRB): the administrative body established by MCC in 2013 to (i) develop, review and approve guidelines and procedures (including modifications thereto) for data activities; (ii) review and approve proposals related to data disclosure; and (iii) notify the MCC Incident Response team in the event of an identified, specific disclosure risk (spill, breach, etc.) and follow MCC protocol for risk management.

Evaluation Design Report (EDR): standard contract deliverable for independent evaluation contractors where the evaluation design is fully documented and approved by the relevant MCC Evaluation Management Committee

HHS: the United States Department of Health and Human Services.

indirect identifier/quasi-identifier: data that can be used to identify a person (individual or legal) through association with another variable(s). These include unique, observable or other characteristics that may identify a specific data provider (or household, community, school, etc.) even when direct identifiers are removed.

informed consent: action required in research to operationalize respect for persons, where research subjects or data providers are informed of the objectives, duration, and description of the research, its expected benefits and risks, promises of confidentiality, how and who data will be shared with, and that their participation is voluntary.

Institutional Review Board (IRB): an administrative body established to assure that appropriate steps are taken to protect the rights and welfare of humans participating as subjects in research. To accomplish this purpose, IRBs use a group process to review, both in advance and periodically, research protocols and related materials (e.g., informed consent documents and investigator brochures) to ensure protection of the rights and welfare of human subjects of research.

justice: in research refers to the just distribution of the risks and burdens of the research and the benefits expected to be produced by the research.

linkage documentation: documents and other materials unrelated to the applicable data activity but that may support re-identification efforts or at least mitigate de-identification efforts.

MCC staff: as defined in Section 1.2, individuals employed by MCC.

personally identifiable information (PII): information that can be used, on its own or in conjunction with other information that is linked or linkable to a specific individual (or household, community, school, etc.), to determine the identity of a data provider or otherwise locate or contact the data provider. PII includes both direct and indirect (or quasi) identifiers.

p-hacking: known also as "data-mining" or "specification search" defines all the analytical alternatives that a research might test in order to obtain a statistically significant result. Examples include: restrict the sample, test subgroups or redefine variable *after* looking at the final data.

primary data handlers: MCC M&E staff and Contractor Key Personnel

re-identification: any process that restores the association between a set of de-identified data and the data provider.

reporting guidelines: a standardized procedure to report on study design, implementation, analysis, and interpretation of findings.

reproducibility/credibility crisis: general term to describe research findings that describe several problems across scientific fields. These include: low rates of computational reproducibility, high prevalence of publication bias and p-hacking.

research protocol: a tool for documenting the planned research design and practices of a research activity, governing the activity's implementation, and communicating its objectives and expected contributions.

researcher: an individual working for a contractor to lead one or more data activities.

respect for persons: an ethical principle of research that incorporates at least two ideas: (i) individuals are treated as autonomous agents and (ii) individuals with diminished autonomy are entitled to protection. In most cases, respect for persons requires that research subjects or data providers enter into the research voluntarily and with adequate information.

sample frame: the list from which units are drawn for a sample. The 'list' may be an actual listing of units, as in a phone book from which phone numbers will be sampled, or some other description of the population, such as a map from which areas will be sampled.

sample unit: the single value by which an aggregate sample is divided; each sample unit is regarded as individual and indivisible when the selection is made (for example: in an education evaluation, the sample units may be the (i) schools, (ii) teachers, (iii) households, and (iv) children).

sensitive data: information that may pose a risk to the data provider if it is collected or released in a way that is linkable to the data provider (e.g., income, assets or health status).

study registration: A public, brief description of a study before data is available for analysis.

Transparency Statement: the contractor-authored document that documents the extent to which analysis in a published report can or cannot be reproduced with available data and documentation, and justifications for why reproduction cannot be facilitated.

vulnerability: refers to a diminished ability to fully safeguard one's own interest in the context of a specific research project. This may be caused by limited decision-making capacity or limited access to social goods, such as rights, opportunities, and power. Individuals or groups may experience vulnerability to different degrees and at different times, depending on their circumstances.

References

Alderman, Harold, Jishnu Das, and Vijayendra Rao. 2016. "Conducting Ethical Economic Research: Complications from the Field." In *The Oxford Handbook of Professional Economic Ethics* edited by George DeMartino and Deirdre McCloskey. Oxford University Press, April.

Brodeur, Abel, Mathias Lé, Marc Sangnier, and Yanos Zylberberg. 2016. "Star Wars: The Empirics Strike Back." American Economic Journal: Applied Economics, 8 (1): 1-32. DOI: 10.1257/app.20150044

Dupriez, Olivier and Ernie Boyko. 2010. Dissemination of Data Files. Formulating Policies and

Procedures. International Household Survey Network, IHSN Working Paper No 005.

Franco, Annie, Neil Malhotra, and Gabor Simonovits. 2014. "Social Science. Publication Bias in the Social Sciences: Unlocking the File Drawer." Science 345 (6203): 1502–5.

Gertler, Paul, Sebastian Galiani, and Mauricio Romero. 2018. "How to Make Replication the Norm." Nature 554 (7693): 417–19.

Glennerster, Rachel, and Shawn Powers. 2016. "Assessing Risk and Benefit: Ethical Considerations for Running Randomized Evaluations, Especially in Developing Countries." In *The Oxford Handbook of Professional Economic Ethics* edited by George DeMartino and Deirdre McCloskey. Oxford University Press, April.

Hoces de la Guardia, Fernando, and Jennifer Sturdy. February 2019. Best practices in transparent, reproducible, and ethical research. IDB Technical Note 1635. http://dx.doi.org/10.18235/0001564

Hanson, Heather and Catherine Marschner. January 2015. "Transparency" Millennium Challenge Corporation Principles into Practice. https://assets.mcc.gov/reports/paper-2015001163301-principles-transparency.pdf

Leamer, E. E. (1983). Let's take the con out of econometrics. The American Economic Review, 73(1), 31-43.

Miguel, E., C. Camerer, K. Casey, J. Cohen, K. M. Esterling, A. Gerber, R. Glennerster, et al. 2014. "Promoting Transparency in Social Science Research." *Science* 343 (6166): 30–31. doi:10.1126/science.1245317.

NISTIR 8053, De-Identification of Personal Data, Simson Garfinkel, September 2015, National Institute of Standards and Technology, Gaithersburg, MD. http://dx.doi.org/10.6028/NIST.IR.8053.

NIST 2016, De-Identification of Government Data
http://csrc.nist.gov/publications/PubsDrafts.html#SP-800-188

Poe, Ted. (10/20/2015). “H.R.3766 – Foreign Aid Transparency and Accountability Act of 2016.” Legislation. July 15. https://www.congress.gov/bill/114th-congress/house-bill/3766.

Ryan, Paul. 2016. “H.R.1831 – 114th Congress (2015-2016): Evidence-Based Policymaking Commission Act of 2016.” Legislation. March 30. https://www.congress.gov/bill/114th-congress/house-bill/1831.

Sturdy, Jennifer, Sixto Aquino, and Jack Molyneaux. 2014. “Learning from Evaluation at the Millennium Challenge Corporation”. *Journal of Development Effectiveness*. Taylor & Francis. DOI:10.1080/19439342.2014.975424

Annexes

These supporting documents can help evaluators produce deliverables promoting transparency, reproducibility, and ethics in line with the TREDD Guidelines:

1. Informed Consent Template
2. Metadata Template and Instructions
3. Evaluator Report Templates
4. Satellite Imagery Guidance
5. Data Package Worksheet
6. Transparency Statement Template
7. DRB Charter
8. DRB Review Process (Visual)
9. Identifiable Data Sharing Form
10. Disclosure Incident Form

Endnotes

1. The Common Rule was published in 1991 and revised in January 2017; it is codified in separate regulations by 15 Federal departments and agencies (published here – https://www.hhs.gov/ohrp/regulations-and-policy/regulations/finalized-revisions-common-rule/index.html.) The Common Rule's focus is safeguarding the rights and welfare of human subjects involved in federally funded research. Although MCC is not a signatory of the Common Rule, MCC recognizes and acknowledges the ethical and privacy implications involved in research involving human subjects.
2. These tasks are also built into the standard scope of work and personnel qualifications for MCC-funded contractors.
3. Other public platforms may be identified for data activities outside the principal scope of these guidelines.
4. To date, this has mostly related to independent evaluation-related data, but may also include economic analysis surveys, due diligence studies, and other studies informing operations.
5. These guidelines may be revised and updated from time to time, and such revision will be promptly posted on the MCC website. If the guidelines are updated during the course of an evaluation or contract term, staff and contractors should apply the most recent, approved version to their work to the extent possible.
6. Information available at https://strategy.data.gov/.
7. Descriptions and key references adapted from Hoces de la Guardia and Sturdy (2018).
8. MCC is aligning its requirements with the social science research community by requesting renewal every 4 years, or in the event of a major change to the regulations. See for example the requirements of Harvard University, Georgetown University, and Stanford University, each of which requires renewal every three years.
9. The Freedom of Information Act (5 U.S.C. 552 (1996)) was first passed in 1967 and gives the public a statutory right of access to federal agency records with the aim of encouraging government accountability through transparency. Like all federal agencies of the US Government, MCC may withhold information pursuant to the exemptions and exclusions to disclosure contained in the statute. See MCC's FOIA Regulation (available at https://www.mcc.gov/resources/doc/mcc-foia-regulation-2018) for a detailed description of the rules MCC follows in processing requests for records under the Freedom of Information Act.
10. https://www.fda.gov/RegulatoryInformation/Guidances/ucm126420.htm
11. This is an **explicit requirement** in the Revised Common Rule (45 CFR 46.116(b)(9), subpart A).
12. This is an **option** under the Revised Common Rule (45 CFR 46.116(b)(9), subpart A).
13. In 1998, Congress amended the Rehabilitation Act of 1973 to require Federal agencies to make their electronic and information technology (EIT) accessible to people with disabilities. The law (29 U.S.C. § 794 (d)) applies to all Federal agencies when they develop, procure, maintain, or use electronic and information technology. Under Section 508 of the Act, agencies must give disabled employees and members of the public access to information that is comparable to access available to others. The United States Access Board discusses the Section 508 law and its responsibility for developing accessibility standards for EIT to incorporate into regulations that govern Federal procurement practices. More information is available online at https://www.section508.gov/.
14. https://www.nist.gov/publications/nist-special-publication-800-88-revision-1-guidelines-media-sanitization
15. This can also help contractors meet evolving requirements for journal publications (for example, see AEA data and code submission requirements https://www.aeaweb.org/journals/policies/data-code).

16. Contractors and MCC are encouraged to watch the discussions here on transparency and qualitative research – https://www.qualtd.net/.
17. See https://www.hhs.gov/ohrp/compliance-and-reporting/guidance-on-reporting-incident/index.html.

Reducing Poverty Through Growth

MILLENNIUM
CHALLENGE CORPORATION
UNITED STATES OF AMERICA